Learning-Aided Multi-UAV Online Trajectory Coordination and Resource Allocation for Mobile WSNs

Lu Chen*, Suzhi Bi*, Xiaohui Lin*, Zheyuan Yang⋆, Yuan Wu†, and Qiang Ye‡
*College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China 518060
⋆Department of Information Engineering, The Chinese Unviersity of Hong Kong, Shatin, N. T., Hong Kong SAR
†State Key Lab of Internet of Things for Smart City, The University of Macau, Macao SAR
‡Department of Computer Science, Memorial University of Newfoundland, Canada
E-mail: chenlu2020@email.szu.edu.cn, {bsz,xhlin}@szu.edu.cn, yz019@ie.cuhk.edu.hk, yuanwu@um.edu.mo, qiangy@mun.ca

***Abstract*—In this paper, we consider a multi-UAV enabled wireless sensor network (WSN) where multiple unmanned aerial vehicles (UAVs) gather data from multiple randomly moving sensor nodes (SNs). We aim to minimize the long-term average energy consumption of all SNs while satisfying their average data rate requirements and energy constraints of the UAVs. We solve the problem by jointly optimizing the multi-UAV's trajectories, communication scheduling and SN's association decisions. In particular, we formulate it as a multi-stage stochastic mixed integer non-linear programming (MINLP) problem and design an online algorithm that integrates Lyapunov optimization and deep reinforcement learning (DRL) methods. Specifically, we first decouple the original multi-stage stochastic MINLP problem into a series of per-slot deterministic MINLP subproblems by applying Lyapunov optimization. For each per-slot problem, we use model-free DRL to obtain the optimal integer UAV-SN associations and model-based method to optimize the UAVs' trajectories and resource allocation. Simulation results reveal that although the communication environments change stochastically and rapidly, our proposed online algorithm can produce real-time solution that achieves high system performance and satisfies all the constraints.**

I. Introduction

Due to its deployment flexibility, low cost, and formidible data collection ability, wireless sensor network (WSN) has been widely applied in battlefield surveillance, environmental monitoring and target tracking [1]. Generally, a WSN consists of a lot of sensor nodes (SNs) powered by batteries, which are inconvenient and costly to replace or recharge. Because the distance between SNs and the data collection center can vary significantly in a WSN, the energy consumed on data transmission may differ greatly, which can result in early energy depletion for some sensors. The inherent battery constraint unfairness problem has largely reduced the lifespan of WSNs. Meanwhile, the maturity of unmanned aerial vehicle (UAV) technology and the ever-decreasing manufacturing cost have made the proliferations of UAV in civil and military areas [2] - [4]. Using UAV as a mobile sink to collect sensing data in WSNs can effectively mitigate the above mentioned problems in conventional WSNs. Specifically, the UAVs can fly closely to the SNs to reduce the energy consumed on data transmission of SNs and prolong the lifetime of WSN [5]. Beside the consideration of energy consumptions of SNs on communications, it is also important to design UAV trajectories to reduce energy consumed on UAV propulsion for sustainable UAV operations.

The joint optimization of trajectory and communication resource allocation to achieve optimum energy efficiency has been extensively studied. Zeng et al. [6] consider dispatching a single UAV to communicate with multiple fixed SNs and focus on minimizing the energy consumption of UAV under the communication throughput constraints through trajectory optimization. Given the UAV's energy budget, [7] minimizes the maximal energy consumption among all SNs. Zhan and Huang [8] present a performance tradeoff between propulsion energy of UAV and communication energy of SN. In particular, they minimize the weighted sum energy consumption of UAV and SNs by jointly optimizing UAV trajectory, mission time and the wake-up scheduling. Unlike in [6] - [8] where a single UAV is used to collect data, in [9], the authors consider using multiple UAVs to collect data from fixed SNs in a "flying and hovering" manner. By jointly optimizing the UAV trajectory, wake-up scheduling and UAV-SN association, the mission completion time and the energy consumption of the SNs are minimized.

Most of the exiting works assume that users' locations are fixed, and adopt offline algorithms to optimize the UAV's trajectory and resource allocation. In fact, in many scenarios such as battlefield and mobile games, the locations of users and the associated real-time transmission requirements can vary frequently. Thus, an online real-time algorithm is needed to optimize both the UAV's trajectory and resource allocation. In [10], Yang et al. consider a single-UAV enabled MEC system to serve users of stochastic movements. They design online UAV trajectory using a Lyapunov optimization method to minimize the weighted sum average energy consumption of ground mobile users under UAV average energy consumption constraint. The recent work in [11] considers a multi-UAV assisted downlink wireless network in which the ground terminals are mobile. It designs 3D trajectories of the multi-UAVs using a constrained Deep Q-Network to maximize the real-time downlink capacity while guaranteeing that all ground terminals are served within a deadline. The energy consumption of the UAVs, however, is not the major design concern. To serve mobile users with multiple energy-constrained UAVs, both the trajectories and resource allocation should be coordinated to minimize the energy consumption of both UAVs and SNs while satisfying the system performance requirements, which requires fast online algorithms.

In this work, we employ multi-UAVs to undertake data collection from mobile SNs of unknown mobility pattern. We aim to design an online algorithm to minimize the weighted sum transmission energy consumption of the SNs under the average UAV energy consumption and SN transmission rate constraints. This requires online decision of the UAV trajectory, UAV-SN communication association, and communication resource allocation. We formulate it as a multi-stage stochastic mixed integer non-linear programming (MINLP) problem. To cope with it, we adopt an integrated learning and optimization framework. Firstly, we decouple the original problem into per-stage deterministic MINLP subproblems by using the Lyapunov optimization. Then, we solve each per-slot subproblem with a proposed low-complexity algorithm. Specifically, we use model-free DRL to tackle the hard combinatorial UAV-SN association problem and model-based optimization to control the trajectory and transmission time allocation for each UAV respectively. Simulation results reveal that, the online algorithm not only quickly generates high-quality solution, but also satisfies all the long-term performance requirements.

II. System Model and Problem Formulation

We consider a multi-UAV enabled WSN consisting of M UAVs and K mobile ground SNs. We denote $\mathcal{M} = \{1, 2, ..., M\}$ and $\mathcal{K} = \{1, 2, ..., K\}$ as the sets of UAVs and SNs, where $K \geq M$. The UAVs collect data from the ground users in duration T with limited on-board battery energy. For ease of illustration, we discretize duration T into a set of N time slots equally denoted by $\mathcal{N} = \{1, 2, ..., N\}$. For each time slot, its length $\Delta = T/N$ is chosen to be sufficiently small, thus the positions of UAVs and users can be considered to be unchanged in this slot [12].

We suppose that users follow the Gauss-Markov mobility model [13]. Specifically, the velocity of user k $(s_k, k \in \mathcal{K})$ at time slot $n+1$ is $\mathbf{v}_k[n+1] = \alpha\mathbf{v}_k[n] + (1-\alpha)\bar{\mathbf{v}}_k + \bar{\sigma}\sqrt{1-\alpha^2}\mathbf{w}_k[n]$, where the vector $\mathbf{w}_k[n]$ denotes an independent random Gaussian distribution of zero-mean and variance σ^2. $\alpha \leq 1$ denotes memory level. $\bar{\mathbf{v}}_k$ and $\bar{\sigma}$ are parameters denoting the average speed and asymptotic standard deviation of velocity, respectively. Accordingly, the location of s_k at time slot $n+1$ is $\mathbf{p}_k[n+1] = \mathbf{p}_k[n] + \mathbf{v}_k[n]\Delta$. We suppose that UAVs know the locations of the users at current time slot (e.g., via location report) but are uncertain of the users' future traces.

We assume that all UAVs start from their initial locations and finally reach their predetermined destinations after a mission period T. Each UAV flies at a fixed height h and gathers data from ground users constrained by a maximum speed V_{max}. The trajectory of UAV m $(u_m, m \in \mathcal{M})$ at time slot n is $\mathbf{q}_m[n]$ with $\mathbf{q}_m[1] = \mathbf{q}_{I,m}$ and $\mathbf{q}_m[N+1] = \mathbf{q}_{F,m}$ representing the u_m's initial and final position respectively.

In order to avoid the interference among UAVs, we suppose that all UAVs collect data using orthogonal frequency channels, where each is allocated with a non-overlapping bandwidth WHz. Meanwhile, each UAV receives data from its associated users via time division multiple access (TDMA). Specifically, for u_m, time slot n is further partitioned into K sub-slots $\{\delta_{m,k}[n]\}_{k=1}^{K}$ with $\sum_{k=1}^{K} \delta_{m,k}[n] \leq \Delta$, and sub-slot $\delta_{m,k}[n]$ is the transmission time assigned to s_k. To characterize the association of s_k with u_m, we define a binary variable $x_{m,k}[n]$, where $x_{m,k}[n] = 1$ indicates that s_k associates with u_m at time slot n and $x_{m,k}[n] = 0$ otherwise. Furthermore, it's assumed that each user is served by at most one UAV at time slot n.

We characterize the UAV-SN links by probabilistic line-of-sight (LoS) channel model [14]. Specifically, the LoS probability of s_k at time slot n is approximated to $\mathbb{P}(LoS, \rho_{m,k}[n]) = 1/(1 + ae^{-b(\rho_{m,k}[n]-a)})$, where a and b are parameters related to environment, and $\rho_{m,k}[n] = \frac{180}{\pi}\arctan(\frac{h}{\|\mathbf{q}_m[n]-\mathbf{p}_k[n]\|})$. Furthermore, we can obtain the expected channel power gain as $g_{m,k}[n] = \frac{\hat{\mathbb{P}}(LoS,\rho_{m,k}[n])g_0}{(h^2+\|\mathbf{q}_m[n]-\mathbf{p}_k[n]\|^2)^{\frac{\tilde{\iota}}{2}}}$, where $\hat{\mathbb{P}}(LoS, \rho_{m,k}[n]) = \mathbb{P}(LoS, \rho_{m,k}[n]) + (1 - \mathbb{P}(LoS, \rho_{m,k}[n]))\kappa$ is the regularized LoS probability, $\kappa < 1$ represents the attenuation factor due to the non-line-of-sight (NLoS) channel propagation, $\tilde{\iota}$ is the path loss exponent, g_0 is the channel gain at the reference distance $d_0 = 1$ m. Accordingly, the achievable uplink rate from s_k to u_m can be expressed as $R_{m,k} = W\log_2(1 + \frac{\gamma_{m,k}[n]}{(h^2+\|\mathbf{q}_m[n]-\mathbf{p}_k[n]\|^2)^{\iota}})$, where $\gamma_{m,k}[n] = \frac{P_k\hat{\mathbb{P}}(LoS,\rho_{m,k}[n])g_0}{N_0}$, $\iota = \frac{\tilde{\iota}}{2}$, P_k is the fixed transmission power of s_k and N_0 is the noise power. Therefore, we derive the expected transmission data rate of s_k from $R_k[n] = \sum_{m=1}^{M} x_{m,k}[n]R_{m,k}[n]\delta_{m,k}[n]$. Accordingly, the energy consumption at interval n of user s_k is $E_k[n] = \sum_{m=1}^{M} x_{m,k}[n]P_k\delta_{m,k}[n]$.

In this paper, we ignore the communication energy and adopt the existing mathematical model [6] to characterize the propulsion energy of rotary-wing aircraft as $P_m(v_m[n])[n] = C_1(1 + 3\frac{v_m^2[n]}{v_{tip}^2}) + C_2\sqrt{\sqrt{C_3 + \frac{v_m^4[n]}{4}} - \frac{v_m^2[n]}{2}} + C_4 v_m^3[n]$, where $v_m[n]$ is the velocity of u_m during interval n, v_{tip} represents the tip speed of the rotor, and C_1, C_2, C_3, C_4 are constants depending on the UAV's weight and its aerodynamic parameters [15]. During the time slot n, propulsion energy of u_m is $E_m[n] = P_m[n]\Delta$. And $v_m[n] = \frac{\|\mathbf{q}_m[n+1]-\mathbf{q}_m[n]\|}{\Delta}$.

In our work, we aim to design an online algorithm to minimize the long-term sum energy consumption of all SNs under average energy constraints of UAVs, and average data requirements of SNs. Specifically, we jointly optimize the UAV trajectories and the communication resource allocation by solving the following multi-stage stochastic optimization problem.

$$\text{(P1)}: \min_{\mathbf{q}[n],\mathbf{x}[n],\delta[n]} \lim_{N\to\infty} 1/N\sum_{n=1}^{N}\sum_{k=1}^{K} w_k E_k[n], \tag{1a}$$

$$s.t. \lim_{N\to\infty} 1/N\sum_{n=1}^{N} R_k[n] \geq \beta_k, \forall k, \tag{1b}$$

$$\lim_{N\to\infty} 1/N\sum_{n=1}^{N} \mathbb{E}\left\{E_m[n]\right\} \leq e_m, \forall m, \tag{1c}$$

$$\sum_{m=1}^{M} x_{m,k}[n] \leq 1, \forall k, n, \tag{1d}$$

$$x_{m,k}[n] \in \{0, 1\}, \forall m, k, n, \tag{1e}$$

$$\sum_{k=1}^{K} \delta_{m,k}[n] \leq \Delta, \delta_{m,k}[n] \geq 0, \forall m, k, n, \tag{1f}$$

$$\delta_{m,k}[n] \leq x_{m,k}[n]\Delta, \forall m, k, n, \tag{1g}$$

$$\|\mathbf{q}_m[n+1] - \mathbf{q}_m[n]\| \leq V_{max}\Delta, \forall m, n, \tag{1h}$$

$$\|\mathbf{q}_{F,m} - \mathbf{q}_m[n+1]\| \le V_{max}(N-n)\Delta, \forall m, n, \tag{1i}$$

where w_k represents the fix weight of s_k. The variable $\mathbf{q}[n]$ denotes the location vector of UAVs at time slot n, i.e., $\mathbf{q}[n] = \{\mathbf{q}_m[n]\}$. $\mathbf{x}[n]$ consists of UAV-SN associations, that is, $\mathbf{x}[n] = \{\mathbf{x}_{m,k}[n]\}$. $\delta[n]$ represents transmission time allocation between SN and UAVs, i.e., $\{\delta_{m,k}[n]\}$. (1b) represents the data requirements of SNs. (1c) is the propulsion energy constraints of UAVs. (1d) and (1e) are the association constraints. (1f) and (1g) are the constraints on data transmission time. (1h) - (1i) are the trajectory and speed constraints of the UAVs. Due to randomness of the mobile users and uncertainty of channel conditions, it is hard to satisfy the long-term constraints (1b) and (1c) when making online decision without knowing the SNs' future locations. In the following, we adopt an integrated learning and optimization method to solve (P1).

III. Lyapunov-Based Decoupling of (P1)

In this section, we transform the multi-stage stochastic problem (P1) to per-slot deterministic problems for online implementation. To cope with the long-term average constraints (1b) and (1c), we introduce K virtual data queues $\mathbf{Y}[n](i.e., \{Y_k[n]\}_{k=1}^{K})$ and M virtual energy queues $\mathbf{Q}[n](i.e., \{Q_m[n]\}_{m=1}^{M})$. By setting $Y_k[1] = 0$ and $Q_m[1] = 0$, the virtual queues are updated as

$$Y_k[n+1] = \max\{Y_k[n] + \beta_k - R_k[n], 0\}, \forall n \in \mathcal{N}, \tag{2}$$

$$Q_m[n+1] = \max\{Q_m[n] + E_m[n] - e_m, 0\}, \forall n \in \mathcal{N}. \tag{3}$$

Here, we denote $\mathbf{Z}[n] = \{\mathbf{Q}[n], \mathbf{Y}[n]\}$ as the system queue backlogs. Then, the Lyapunov function $L(\mathbf{Z}[n])$ and Lyapunov drift [16] are given by

$$L(\mathbf{Z}[n]) = 1/2(\textstyle\sum_{k=1}^{K} Y_k^2[n] + \sum_{m=1}^{M} Q_m^2[n]), \tag{4}$$

$$\Delta L(\mathbf{Z}[n]) = \mathbb{E}\{L(\mathbf{Z}[n+1]) - L(\mathbf{Z}[n])\}. \tag{5}$$

The corresponding drift-plus-penalty function is [17]:

$$D(\mathbf{Z}[n]) = \Delta L(\mathbf{Z}[n]) + V\mathbb{E}\{E_s[n]|\mathbf{Z}[n]\}, \tag{6}$$

where $E_s[n] = \sum_{k=1}^{K} w_k E_k[n]$ denotes the weighted sum energy consumption. The parameter V is for balancing the system energy consumption and queue stability. To minimize $E_s[n]$ while achieving stability of queue $\mathbf{Z}[n]$, we minimize the $D(\mathbf{Z}[n])$ in per slot.

Proposition 1: For queue backlog $\mathbf{Z}[n]$ at any time slot n, the drift-plus-penalty $D(\mathbf{Z}[n])$ is upper bounded by

$$\begin{aligned} D(\mathbf{Z}[n]) \le C + \textstyle\sum_{k=1}^{K} Y_k[n]\mathbb{E}\{\beta_k - R_k[n]|\mathbf{Z}[n]\} + \\ \textstyle\sum_{m=1}^{M} Q_m[n]\mathbb{E}\{E_m[n] - e_m|\mathbf{Z}[n]\} + V\mathbb{E}\{E_s[n]|\mathbf{Z}[n]\}, \end{aligned} \tag{7}$$

here C is a finite constant determined by the maximum second order moment of $R_k[n]$ and $E_m[n]$.

Proof: The proof is omitted here due to the page limit. ■

Next, we minimize the upper bound of $D(\mathbf{Z}[n])$ opportunistically given $\mathbf{Z}[n]$. After removing the constant term in the right-hand-side (RHS) of (7), we solve in each time slot n to obtain the updated UAV location $\mathbf{q}[n+1]$, communication time allocation $\delta[n]$ and the UVA-SN association $\mathbf{x}[n]$. For simplicity of exposition, we drop the index n, and replace $\mathbf{q}[n+1]$ with $\mathbf{q}'$. Then, the per-slot problem is formulated as

$$\text{(P2)}: \underset{\mathbf{q}',\mathbf{x},\delta}{min} \textstyle\sum_{m=1}^{M} Q_m E_m - \sum_{k=1}^{K} Y_k R_k + V E_s, \tag{8a}$$

$$s.t. \textstyle\sum_{m=1}^{M} x_{m,k} \le 1, \forall k, \tag{8b}$$

$$x_{m,k} \in \{0,1\}, \forall m, k, \tag{8c}$$

$$\textstyle\sum_{k=1}^{K} \delta_{m,k} \le \Delta, \forall m, \delta_{m,k} \ge 0, \forall m, k, \tag{8d}$$

$$\delta_{m,k} \le x_{m,k}\Delta, \forall m, k, \tag{8e}$$

$$\|{\mathbf{q}'}_m - \mathbf{q}_m\| \le V_{max}\Delta, \forall m, \tag{8f}$$

$$\|\mathbf{q}_m[N+1] - {\mathbf{q}'}_m\| \le V_{max}(N-n)\Delta, \forall m. \tag{8g}$$

Although the multi-stage stochastic problem (P1) is reduced to per-slot deterministic problem (P2), it is still challenging to solve the combinatorial MINLP in each short time slot. In section IV, we will propose an algorithm based on DRL to solve (P2) efficiently.

IV. Online Joint Optimization of Resource Allocation and UAV Trajectory

At the beginning of the nth time slot, we make an observation of $\eta^n = \{\{g^n_{m,k}\}^{M,K}_{m=1,k=1}, \{Q^n_m\}^{M}_{m=1}, \{Y^n_k\}^{K}_{k=1}\}$ and take an action $\{\mathbf{x}^n, \mathbf{y}^n\}$, where $\mathbf{y}^n = \{\mathbf{q}^n, \delta^n\}$. Furthermore, we find that given the discrete variable $\mathbf{x}^n$ in (P2), it can be transformed into a convex problem by applying the successive convex approximation (SCA) technique. We will present the method to optimize $\mathbf{y}^n$ when $\mathbf{x}^n$ is given in Algorithm 2. For now, we use $G(\mathbf{x}^n, \eta^n)$ to represent the objective value of (P2) after optimizing $\mathbf{y}^n$ given $\mathbf{x}^n$. Thus, (P2) is converted to seeking the optimal UAV-SN association $(\mathbf{x}^n)^*$ as shown below:

$$\text{(P3)}: (\mathbf{x}^n)^* = \arg \min_{\mathbf{x}^n \in \{0,1\}^{M \times K}} G(\mathbf{x}^n, \eta^n), \tag{9}$$

To solve (P3) optimally, it in general requires searching over all the combinations of $\mathbf{x}^n$, which is computationally prohibitive in practice. To achieve real-time control, we apply an integrated DRL and optimization method to solve (P3). As shown in Fig.1, the frame can be divided into four modules introduced as follows.

1) Actor Module: The actor module is based on a deep neural network parameterized by the coefficient θ^n in the nth time slot. In each time slot, we input η^n into DNN's input layer and obtain a set of relaxed UAV-SN association options $\hat{\mathbf{x}}^n$ at the output layer of DNN. Here, we use a sigmoid function to activate the output layer. The relationship between input and output layers is defined as:

$$\Pi_{\theta^n} : \eta^n \mapsto \hat{\mathbf{x}}^n = \left\{\hat{x}^n_{m,k} \in [0,1], m \in \mathcal{M}, k \in \mathcal{K}\right\}, \tag{10}$$

We denote $\hat{\mathbf{x}}^n_k = \left[\hat{x}^n_{1,k}, ..., \hat{x}^n_{m,k}, ..., \hat{x}^n_{M,k}\right]$ as the relaxed association of s_k. Then, we quantize $\hat{\mathbf{x}}^n$ into A candidate integer associations denoted as

$$\Upsilon_A : \hat{\mathbf{x}}^n \mapsto \Omega^n = \left\{\mathbf{x}^n_j \mid j = 1, ..., A\right\}, \tag{11}$$

where $\mathbf{x}^n_j = \{x^n_{j,m,k} \mid x^n_{j,m,k} \in \{0,1\}\}$ denotes the jth candidate. We denote $\mathbf{x}^n_{j,k} = \left[x^n_{j,1,k}, ..., x^n_{j,m,k}, ..., x^n_{j,M,k}\right]$ as the jth candidate association of s_k, such that $\mathbf{x}^n_j = \{\mathbf{x}^n_{j,k}\}$. Compared with quantizing $\hat{\mathbf{x}}^n$ to its nearest binary vector, the diversity in the candidate set prevents the DNN from converging to a sub-optimal result prematurely during training. Specifically, we obtain $\mathbf{x}^n_1$ by quantizing each $\hat{\mathbf{x}}^n_k$ into $\mathbf{x}^n_{1,k}$ with the following rule:

$$x^n_{1,m,k} = \begin{cases} 1 & \text{if } \hat{x}^n_{m,k} \text{ is the maxmium among } \hat{\mathbf{x}}^n_k \\ 0 & \text{othervise.} \end{cases} \tag{12}$$

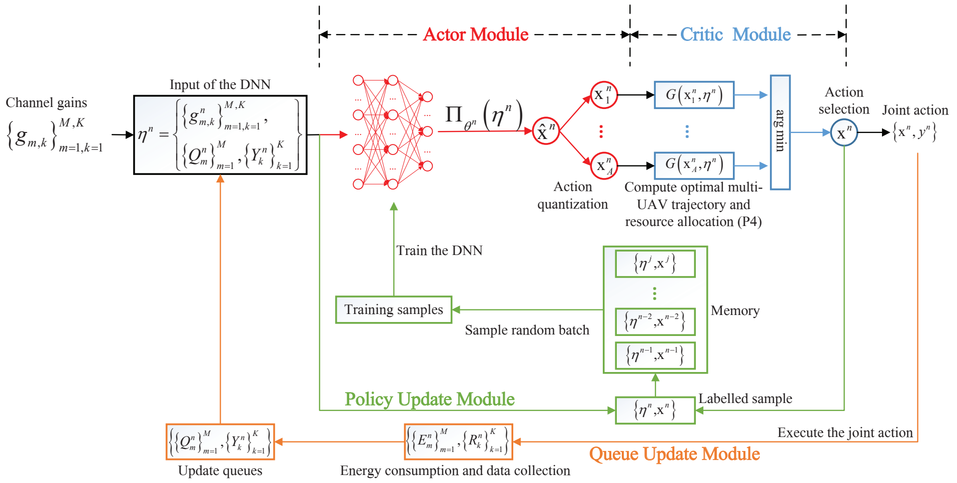


Fig. 1: The schematics of the proposed algorithm.

for $m = 1, \cdots, M$ and $k = 1, \cdots, K$. To obtain the remaining actions, we first add independent identically distributed noise $\mathbf{n}_j$ to $\hat{\mathbf{x}}^n$ for $A-1$ times to obtain a set of relaxed continuous associations $\{\tilde{\mathbf{x}}_j^n, j = 2, ..., A\}$, where $\mathbf{n} \sim \mathcal{N}(0, \mathbf{I}_N)$ with $\mathbf{I}_N$ being an identity matrix. Then, we normalize $\{\tilde{\mathbf{x}}_j^n\}$ by using the sigmoid function to bound the values within [0, 1], denoted by $\tilde{\mathbf{x}}_j^n = sigmoid(\hat{\mathbf{x}}^n + \mathbf{n}_j)$. Let $\tilde{\mathbf{x}}_{j,k}^n = [x_{j,1,k}^n, ..., x_{j,m,k}^n, ..., x_{j,M,k}^n]$. We obtain the remaining actions $\{\mathbf{x}_2^n, ..., \mathbf{x}_A^n\}$ by quantizing each $\tilde{\mathbf{x}}_j^n$ as

$$x_{j,m,k}^n = \begin{cases} 1 & \text{if } \tilde{x}_{j,m,k}^n \text{ is the maxmium among } \tilde{\mathbf{x}}_{j,k}^n \\ 0 & \text{otherise.} \end{cases} \tag{13}$$

Now, we obtain all the A candidate quantization associations.

2) Critic Module: Given the quantized candidate set Ω^n in (11), the critic module uses the model information to evaluate the candidates by solving a joint optimization of multi-UAV trajectory and resource allocation. In particular, the selected optimal association $\mathbf{x}^n$ is denoted by

$$\mathbf{x}^n = \arg \min_{\mathbf{x}_j^n \in \Omega^n} G(\mathbf{x}_j^n, \eta^n). \tag{14}$$

The method to calculate $G(\mathbf{x}_j^n, \eta^n)$ is given in Algorithm 2. The best association decision $\mathbf{x}^n$, along with the corresponding transmission time allocation δ and the UAV location update $\mathbf{q}'$, will be executed to control the UAVs and SNs in the current time slot, i.e., the solution to (P2).

3) Policy Update Module: After obtaining the best action in (14), we store the state-action pair $(\eta^n, \mathbf{x}^n)$ as labeled sample in a finite memory. Here, we only store the latest z data and DNN training begins after gathering $z/2$ samples. Afterwards, we train the DNN every τ time slots to prevent model from over-fitting. In every training session, we first pick a batch of labeled samples from the memory randomly, and update θ^n by minimizing the average cross-entropy loss function over the data samples. When the training is finished, we update the DNN parameter in next slot to θ^{n+1}.

4) Queue Update Module: In the critic module, and we can derive the propulsion energy consumption $\{E_m[n]\}_{m=1}^M$ and the data gathered from their connected SNs $\{R_k[n]\}_{k=1}^K$. We then update the energy and data queues as $\{Q_m[n]\}_{m=1}^M$ and $\{Y_k[n]\}_{k=1}^K$ using (2) and (3) at the beginning of $(n+1)$th slot, respectively. This also updates η^{n+1} as the input to the actor module in a new iteration.

Within such a loop, the DNN learns continually from the latest d labeled samples to improve its policy that achieves a better performance by solving (P3). We summarize the pseudo-code in Algorithm 1. The computation complexity is on solving $G(\mathbf{x}_j^n, \eta^n)$ for A times in (14).

Algorithm 1: The proposed learning-aided online algorithm for solving (P1).

```
input  : $M$, $K$, $\{e_m\}_{m=1}^M$, $\{\beta_k\}_{k=1}^K$, $\{w_k\}_{k=1}^K$, $V$, $A$, training periods
         $\Gamma$, training interval $\delta_T$;
output : Control Actions $\{\mathbf{x}^n, \mathbf{y}^n\}_{n=1}^N$;
Initialize $\theta^1$ randomly and memory, $A \leftarrow K$, training time $t \leftarrow 0$, $V_{max}$;
for $\tau = 1, ..., \Gamma$ do
    Initialize $\{Y_k[1] = 0, \mathbf{p}_k[1]\}_{k=1}^K$, $\{Q_m[1] = 0, \mathbf{q}_{I,m}, \mathbf{q}_{F,m}\}_{m=1}^M$;
    for $n = 1, ..., N$ do
        Observe the DNN input $\eta^n$;
        Generate a relaxed association $\hat{\mathbf{x}}^n = \Pi_{\theta^n}(\eta^n)$;
        Quantize $\hat{\mathbf{x}}^n$ into $A$ binary associations;
        Compute $G(\mathbf{x}_j^n, \eta^n)$ with $\mathbf{x}_j^n$ using algorithm 2;
        Select the optimal $\mathbf{x}^n$ and execute the joint action $(\mathbf{x}^n, \mathbf{y}^n)$;
        Update the memory by adding $(\eta^n, \mathbf{x}^n)$;
        if $mod(t, \delta_T)$=0 then
            Randomly pick a batch of data set $\mathcal{S}^n$from the memory;
            Train the DNN with $\mathcal{S}^n$ and update $\theta^n$ using the Adam
              algorithm;
        end
        $n \leftarrow n + 1$;
        Update $\{Q_m[n]\}_{m=1}^M$ and $\{Y_k[n]\}_{k=1}^K$ based on (2) and (3).
        $t \leftarrow t + 1$;
    end
    $\tau \leftarrow \tau + 1$;
end
```

Next, we present how to compute $G(\mathbf{x}_j^n, \eta^n)$ efficiently. Given $\mathbf{x}$, we solve a joint optimization of multi-UAV trajectory and resource allocation in (P2). It can be observed that for each UAV, the optimization problem to obtain the trajectory and time allocation is independent. Thus, we decompose the problem into M sub-problems one for each UAV. Then, for UAV u_m, we solve the following problem.

$$\begin{aligned} \text{(P4)} : & \min_{\mathbf{q}', \delta} V \textstyle\sum_{k=1}^K w_k P_k x_{m,k} \delta_{m,k} + Q_m E_m - \textstyle\sum_{k=1}^K \\ & W \log_2 \Big(1 + \frac{\gamma_{m,k}}{(h^2 + \|\mathbf{q}_m - \mathbf{p}_k\|^2)^\iota}\Big) x_{m,k} \delta_{m,k}, & \text{(15a)} \\ & s.t. \textstyle\sum_{k=1}^K \delta_{m,k} \le \Delta, & \text{(15b)} \\ & \delta_{m,k} \ge 0, \forall k, & \text{(15c)} \end{aligned}$$

$$\delta_{m,k} \le x_{m,k}\Delta, \forall k, \tag{15d}$$

$$\left\|\mathbf{q'}_m - \mathbf{q}_m\right\| \le V_{max}\Delta, \tag{15e}$$

$$\left\|\mathbf{q}_m[N+1] - \mathbf{q'}_m\right\| \le V_{max}(N-n)\Delta. \tag{15f}$$

Although the problem is non-convex, it can be efficiently tackled by SCA technique. By introducing auxiliary slack variables to deal with the non-convexity of the UAV's propulsion energy function and the rate of s_k in the objective, we solve the following convex approximation of (P4) in the l-th iteration as below:

$$\begin{aligned}
\text{(P5)}: \min_{\mathbf{q'},\delta,y_m,\varphi_{m,k},\phi_{m,k}} \ & V\textstyle\sum_{k=1}^{K} w_k P_k x_{m,k}\delta_{m,k} - \sum_{k=1}^{K} \\
& Y_k\phi_{m,k} + Q_m(C_1(1+3v_m^2/v_{tip}^2 + C_2 y_m + C_4 v_m^3)\Delta, \\
s.t. \ & \varphi_{m,k}^2/(\delta_{m,k}x_{m,k}) \le R_{m,k}^{(l)}\left\{\mathbf{q'}_m\right\}, \ \forall k, \\
& C_3/y_m^2 \le Y_m^{(l)}\left\{\mathbf{q'}_m, y_m\right\}, \\
& \phi_{m,k} \le \Phi^{(l)}\left\{\varphi_{m,k}\right\}, \ \forall k, \\
& (15b) - (15f),
\end{aligned}$$

where y_m is an auxiliary slack variable to deal with the non-convexity of the UAV's propulsion energy function as $y_m^2 \ge \sqrt{C_3 + \frac{v_m^4}{4}} - \frac{v_m^2}{2}$. According to the first-order Taylor expansion, we have $y_m^2 + v_m^2 \ge (y_m^{(l)})^2 + 2y_m^{(l)}(y_m - y_m^{(l)}) - \frac{\left\|\mathbf{q'}_m^{(l)} - \mathbf{q}_m\right\|^2}{\Delta^2} + \frac{2}{\Delta^2}(\mathbf{q'}_m^{(l)} - \mathbf{q}_m)^T(\mathbf{q'}_m - \mathbf{q}_m) \doteq Y_m^{(l)}\left\{\mathbf{q'}_m, y_m\right\}$. Therefore, $y_m^{(l)}$ is defined as

$$y_m^{(l)} \doteq \sqrt{\sqrt{C_3 + \frac{\left\|\mathbf{q'}_m^{(l)} - \mathbf{q}_m\right\|^4}{4\Delta^2}} - \frac{\left\|\mathbf{q'}_m^{(l)} - \mathbf{q}_m\right\|^2}{2\Delta^2}}. \tag{16}$$

Both $\varphi_{m,k}$ and $\phi_{m,k}$ are auxiliary slack variables to deal with the non-convexity of s_k. Firstly, we set $\xi_{m,k} \le W\log_2(1 + \frac{\gamma_{m,k}}{(h^2 + \|\mathbf{q'}_m - \mathbf{p}_k\|^2)^\iota})$ and we have a concave lower bound of the RHS, which is denoted as $R_{m,k}^{(l)}\left\{\mathbf{q'}_m\right\} \doteq W\log_2(1 + \frac{\gamma_{m,k}}{(h^2 + \left\|\mathbf{q'}_m^{(l)} - \mathbf{p}_k\right\|^2)^\iota}) - \lambda_{m,k}(\|\mathbf{q'}_m - \mathbf{p}_k\|^2 - \left\|\mathbf{q'}_m^{(l)} - \mathbf{p}_k\right\|^2)$, where $\lambda_{m,k} = \frac{W(\log_2 e)\gamma_{k,m}\iota}{[\gamma_{m,k} + (h^2 + \left\|\mathbf{q'}_m^{(l)} - \mathbf{p}_k\right\|^2)^\iota](h^2 + \left\|\mathbf{q'}_m^{(l)} - \mathbf{p}_k\right\|^2)}$. Next, we introduce $\varphi_{m,k}$ to represent the transmission data bits of s_k as $\varphi_{m,k}^2 \le Wlog_2(1 + \frac{\gamma_{m,k}}{(h^2 + \|\mathbf{q'}_m - \mathbf{p}_k\|^2)^\iota})x_{m,k}\delta_{m,k}$. To cope with non-convexity with respect to $\varphi_{m,k}$, we introduce $\phi_{m,k} \le \varphi_{m,k}^2$, and use the first-order Taylor expansion. Thus, we have the concave lower bound as $\varphi_{m,k}^2 \ge (\varphi_{m,k}^{(l)})^2 + 2\varphi_{m,k}^{(l)}(\varphi_{m,k} - \varphi_{m,k}^{(l)}) \doteq \Phi^{(l)}\{\varphi_{m,k}\}$, where $\varphi_{m,k}^{(l)}$ is given by

$$\varphi_{m,k}^{(l)} = \sqrt{Wlog_2(1 + \frac{\gamma_{m,k}}{(h^2 + \|\mathbf{q}'_m{}^{(l)} - \mathbf{p}_k\|^2)^\iota})x_{m,k}\delta_{m,k}^{(l)}}. \tag{17}$$

For each UAV u_m, we obtain the trajectory and transmission time allocation by iteratively solving (P5), where in each iteration, we can solve (P5) efficiently with off-the-shelf optimization solver such as CVX [18]. Meanwhile, we take the solution obtained in the l-th iteration as the new local point in $(l+1)$-th iteration. The iteration stops when improvement of objective value is less than threshold ϵ. We summarize the joint optimization of (P4) in Algorithm 2.

Algorithm 2: SCA-based joint optimization for solving (P4).

```
input  : UAVs' association with SNs $\mathbf{x}$, a feasible solution$\{\mathbf{q'}^{(0)}, \delta^{(0)}\}$;
output : UAVs' positions in the next time slot $\mathbf{q'}^{(opt)}$ and transmission time
         allocation $\delta^{(opt)}$;
Initialization: $\epsilon \leftarrow 0.01$;
for $m = 1, ..., M$ do
    $l \leftarrow 0$;
    repeat
        Obtain $y_m^{(l)}$ and $\varphi_{m,k}^{(l)}$ by (16) and (17), respectively;
        Solve (P5), record the optimal values as $\{\mathbf{q}'_m{}^{(opt)}, \delta_m^{(opt)}\}$ and
          use $o^{(l)}$ to represent the objective value;
        Update the local vales $\mathbf{q}'_m{}^{(l+1)} \leftarrow \mathbf{q}'_m{}^{(opt)}$, $\delta_m^{(l+1)} \leftarrow \delta_m^{(opt)}$;
        $l \leftarrow l + 1$;
    until $o^{(i)} - o^{(i-1)} < \epsilon$;
    $m \leftarrow m + 1$;
end
```

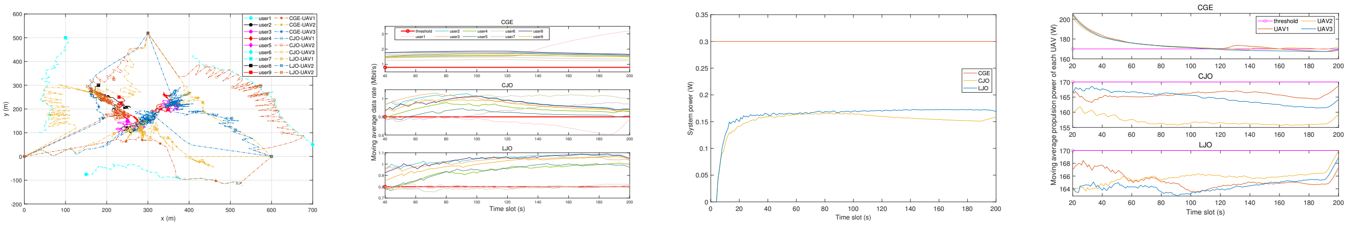
V. Simulation Result

In this section, we evaluate the performance of the proposed algorithms by simulations. We consider a scenario in which three UAVs gather data from nine SNs scattering in a 900m × 800m rectangular area. The UAVs start from their own initial locations and fly to the final locations with $V_{max} = 25$m/s and propulsion energy budget of $e_m = 170$J. Their initial and final locations are $\mathbf{q}_{I,1} = [0,0]$, $\mathbf{q}_{I,2} = [600,0]$, $\mathbf{q}_{I,3} = [300,520]$, $\mathbf{q}_{F,1} = [600,0]$, $\mathbf{q}_{F,2} = [300,520]$, $\mathbf{q}_{F,3} = [0,0]$, respectively. All users start at $\mathbf{p}_1 = [150,75]$, $\mathbf{p}_2 = [250,120]$, $\mathbf{p}_3 = [250,150]$, $\mathbf{p}_4 = [350,200]$, $\mathbf{p}_5 = [370,220]$, $\mathbf{p}_6 = [700,50]$, $\mathbf{p}_7 = [100,500]$, $\mathbf{p}_8 = [180,300]$, $\mathbf{p}_9 = [230,250]$ respectively. Each user has the same initial and average speed. Speed for $s_1 - s_9$ are $[1.8,0]$, $[0.2,0.6]$, $[0.1,0]$, $[0,0.1]$, $[0.6,0.1]$, $[1.2,2.2]$, $[0.6,2.2]$, $[0,0.4]$, $[0.1,0.5]$ separately. And we set the Memory level $\alpha = 0.5$ and standard deviation $\sigma = 2$. We set each mission period as 200s and each time slot Δ as 1s. And communication related parameters are stated as B=1MHz, g_0=-50dB, P_k=0.1W, $N_0 = 10^{-12}$W, $\tilde{\iota}$=2.1, κ=0.2, a=15 and b=0.5. We Suppose that the s_k's data threshold is 0.8 Mbits. The UAV related parameters C_1, C_2, C_3, C_4 are 80, 22, 263.4, 0.0092 respectively. Lyapunov factor is set to V=1200 and each s_k's weight is $1/K$. The DNN actor module adopts a fully-connected multilayer perception, which includes a input layer, two hidden layers and a output layer. The first and the second hidden layers have 256 and 128 neurons respectively. The memory size is 1024 and batch size is 64. We set the training interval $\delta_T = 5$ and the number of candidate association A = 5. For ease of illustration, we refer to our learning-aided joint optimization method as LJO. Moreover, we use the following two algorithms as benchmarks.

Closest Connection + Joint Optimization of Multi-UAV Trajectory and Resource Allocation(CJO): Each SN connects with the closest UAV. Both UAV trajectory and transmission time allocation are controlled by Algorithm 2.

Closest Connection + Geometric Center Tracking + Equal Resource Allocation (CGE): In this benchmark, the SN-UAV association control is the same as CJO. But it is different in UAV trajectory and transmission time allocation. Specifically, each UAV follows the geometric center of its own connected SNs. If a UAV cannot reach the center at the end of current slot, it will fly in the direction towards the center with V_{max}. The SNs connecting to a UAV will be allocated with equal transmission time.

The simulation results are presented in the Fig.2. In Fig.2(a), we present the trajectories obtained from the three methods, and



(a) Users' and UAVs' trajectories

(b) Average SN data rate versus time

(c) System energy power versus time

(d) Average UAV propulsion energy power versus time

Fig. 2: Convergence performance comparisons of different schemes.

we can see that the UAVs in CGE keep following the geometric centers. However, the UAVs in another two algorithms oscillate around their own connected SNs and are inclined to fly closely to the SNs thus to meet the data collection requirements and reduce the energy consumed in data transmission.

The average data, weighted sum energy consumption and propulsion energy consumption at time slot n are denoted as $\frac{1}{n}\sum_{\mu=1}^{n} R_k[\mu]$, $\frac{1}{n}\sum_{\mu=1}^{n}\sum_{k=1}^{K} w_k E_k[\mu]$, $\frac{1}{n}\sum_{\mu=1}^{n} E_m[\mu]$, respectively. In Fig.2(b) and Fig.2(d), it is observed that not all the algorithms satisfy the rate and energy constraints. In particular, in Fig.2(b), we see that CJO fails to meet the rate requirement with user1 transmitting 0.774 Mbits per-slot till the end of the mission, which is less than the requirement of 0.8 Mbits. Besides, we can see from Fig.2(d) that, in CGE, the average propulsion energy consumption of UAV2 is 172.0 J, which exceeds the budget of 170 J. In comparison, LJO satisfies both the data rate requirements of SNs and the energy constraints of UAVs, meanwhile reducing the energy consumption as shown in Fig2.(c). Specifically, the LJO can save 43.7% of energy, compared with the CGE, thanks to the joint optimization of UAV trajectories and transmission time allocation. Notice that the major difference between CJO and LJO is the policy of UAV-SN association. This shows that the simple distance-based UAV-SN allocation method cannot guarantee the long-term performance of the SNs, especially under the highly mobile patterns and dynamic user data rate fluctuations. The proposed method not only quickly generates a high-quality solution, but also guarantees to satisfy the long-term performance requirements.

VI. Conclusion

In this paper, we have proposed a learning-aided online method to optimize the trajectories and resource allocation for WSNs with stochastic mobile SNs. We formulated the problem with long-term objective and performance requirements as a multi-stage stochastic MINLP. The proposed online algorithm integrated Lyapunov optimization and DRL methods to decouple the original multi-stage stochastic MINLP into a series of per-slot deterministic MINLP subproblems and solved each sub-problem with an efficient model-free DRL scheme. Simulation results revealed that the online algorithm could satisfy real-time communication requirements and effectively reduce the energy consumption of SNs.

References

[1] Z. Wei et al., "UAV-Assisted Data Collection for Internet of Things: A Survey," *IEEE Internet of Things Journal*, vol. 9, no. 17, pp. 15460-15483, 1 Sept.1, 2022.

[2] L. Gupta, R. Jain and G. Vaszkun, "Survey of Important Issues in UAV Communication Networks," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1123-1152, Secondquarter 2016.

[3] Y. Zeng, R. Zhang and T. J. Lim, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 36-42, May 2016.

[4] Y. Zeng, Q. Wu and R. Zhang, "Accessing From the Sky: A Tutorial on UAV Communications for 5G and Beyond," *Proceedings of the IEEE*, vol. 107, no. 12, pp. 2327-2375, Dec. 2019.

[5] C. Zhan, Y. Zeng and R. Zhang, "Energy-Efficient Data Collection in UAV Enabled Wireless Sensor Network," *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 328-331, June 2018.

[6] Y. Zeng, J. Xu and R. Zhang, "Energy Minimization for Wireless Communication With Rotary-Wing UAV," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2329-2345, April 2019.

[7] C. Zhan and H. Lai, "Energy Minimization in Internet-of-Things System Based on Rotary-Wing UAV," *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1341-1344, Oct. 2019.

[8] C. Zhan and R. Huang, "Energy Minimization for Data Collection in Wireless Sensor Networks with UAV," *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1-6.

[9] C. Zhan and Y. Zeng, "Completion Time Minimization for Multi-UAV-Enabled Data Collection," *IEEE Transactions on Wireless Communications*, vol. 18, no. 10, pp. 4859-4872, Oct. 2019.

[10] Z. Yang, S. Bi and Y. -J. A. Zhang, "Online Trajectory and Resource Optimization for Stochastic UAV-Enabled MEC Systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 5629-5643, July 2022.

[11] W. Zhang, Q. Wang, X. Liu, Y. Liu and Y. Chen, "Three-Dimension Trajectory Design for Multi-UAV Wireless Network With Deep Reinforcement Learning," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 600-612, Jan. 2021.

[12] Y. Zeng and R. Zhang, "Energy-Efficient UAV Communication With Trajectory Optimization," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3747-3760, June 2017.

[13] S. Batabyal and P. Bhaumik, "Mobility Models, Traces and Impact of Mobility on Opportunistic Routing Algorithms: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1679-1707, thirdquarter 2015.

[14] A. Al-Hourani, S. Kandeepan and S. Lardner, "Optimal LAP Altitude for Maximum Coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569-572, Dec. 2014.

[15] A. Filippone, *Flight Performance of Fixed and Rotary Wing Aircraft*. Amsterdam, The Netherlands: Elsevier, 2006.

[16] M. J. Neely,"Stochastic network optimization with application to communication and queueing systems," *Synthesis Lect. Commun. Netw.*, vol. 3, no. 1, pp. 1–211, Jan. 2010.

[17] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Found. Trends Netw.*, vol. 1, no. 1, pp. 1–144, 2006.

[18] M. Grant and S. Boyd. (Mar. 2014). CVX: MATLAB Software for Disciplined Convex Programming, Version 2.1. [Online]. Available: http://cvxr.com/cvx